wwPDB X-ray Structure Validation Summary Report (i)

Oct 25, 2022 – 05:25 pm BST

PDB ID	:	7R0K
Title	:	Crystal structure of Polymerase I from phage G20c
Authors	:	Welin, M.; Svensson, A.; Hakansson, M.; Al-Karadaghi, S.; Linares-Pasten, J.A.; Jasilionis, A.; Nordberg Karlsson, E.; Ahlqvist, J.
Deposited on	:	2022-02-02
Resolution	:	2.97 Å(reported)

This is a wwPDB X-ray Structure Validation Summary Report for a publicly released PDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (1)) were used in the production of this report:

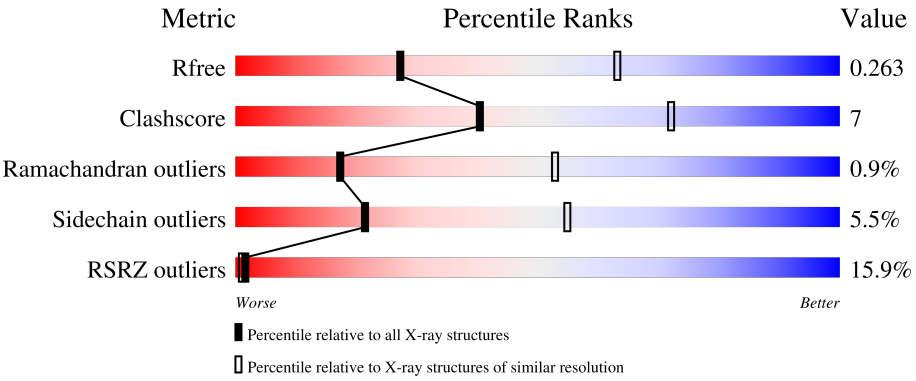
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.31.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0267
CCP4	:	7.1.010 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.31.2

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.97 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	2754 (3.00-2.96)
Clashscore	141614	3103 (3.00-2.96)
Ramachandran outliers	138981	2993 (3.00-2.96)
Sidechain outliers	138945	2996 (3.00-2.96)
RSRZ outliers	127900	2644 (3.00-2.96)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments of the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5% The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	736	11% 77% 18%
1	B	736	19% 73% 17% 8%

2 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 11067 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called DNA polymerase I.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	708	Total	C	N	O	S	0	0	0
			5660	3626	962	1059	13			
1	B	676	Total	C	N	O	S	0	0	0
			5407	3459	922	1013	13			

There are 16 discrepancies between the modelled and reference sequences:

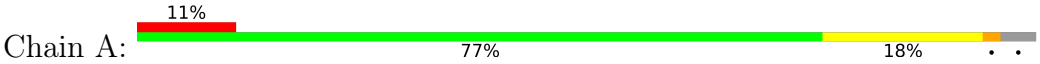
Chain	Residue	Modelled	Actual	Comment	Reference
A	729	LEU	-	expression tag	UNP A0A1L4BKI3
A	730	GLU	-	expression tag	UNP A0A1L4BKI3
A	731	HIS	-	expression tag	UNP A0A1L4BKI3
A	732	HIS	-	expression tag	UNP A0A1L4BKI3
A	733	HIS	-	expression tag	UNP A0A1L4BKI3
A	734	HIS	-	expression tag	UNP A0A1L4BKI3
A	735	HIS	-	expression tag	UNP A0A1L4BKI3
A	736	HIS	-	expression tag	UNP A0A1L4BKI3
B	729	LEU	-	expression tag	UNP A0A1L4BKI3
B	730	GLU	-	expression tag	UNP A0A1L4BKI3
B	731	HIS	-	expression tag	UNP A0A1L4BKI3
B	732	HIS	-	expression tag	UNP A0A1L4BKI3
B	733	HIS	-	expression tag	UNP A0A1L4BKI3
B	734	HIS	-	expression tag	UNP A0A1L4BKI3
B	735	HIS	-	expression tag	UNP A0A1L4BKI3
B	736	HIS	-	expression tag	UNP A0A1L4BKI3

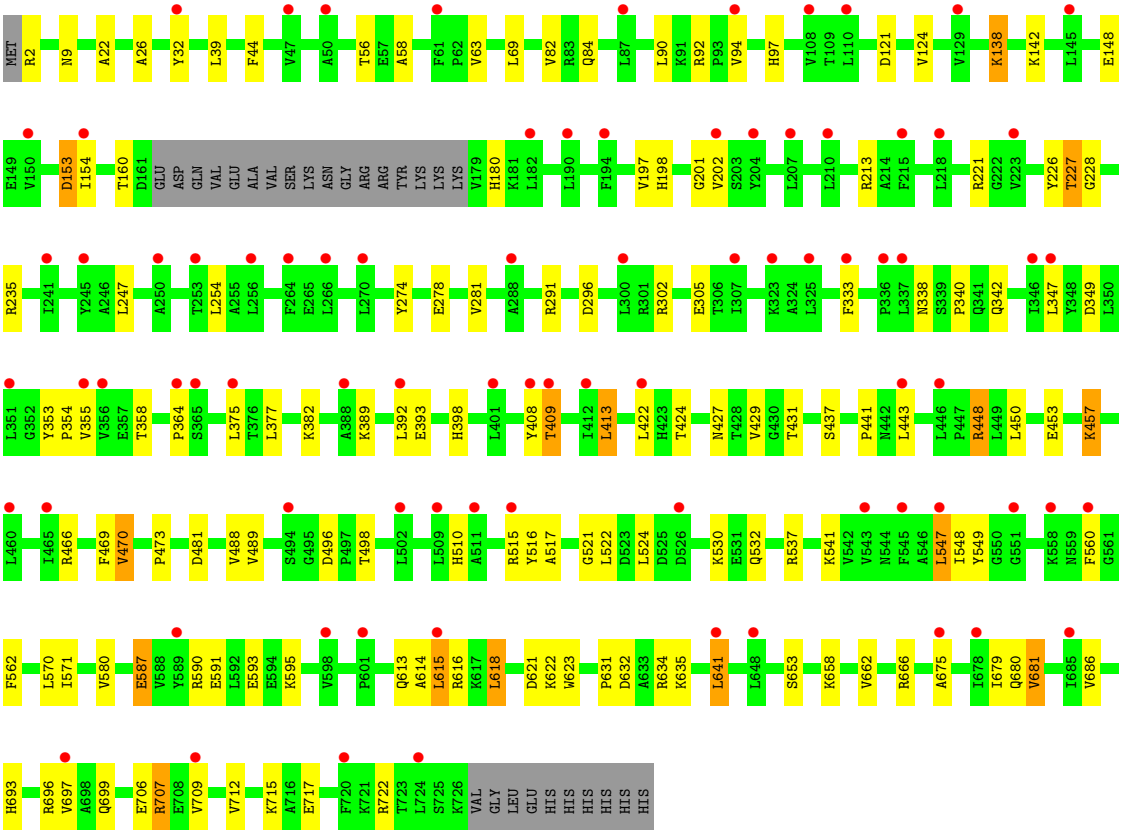
3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density (RSRZ > 2). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

• Molecule 1: DNA polymerase I

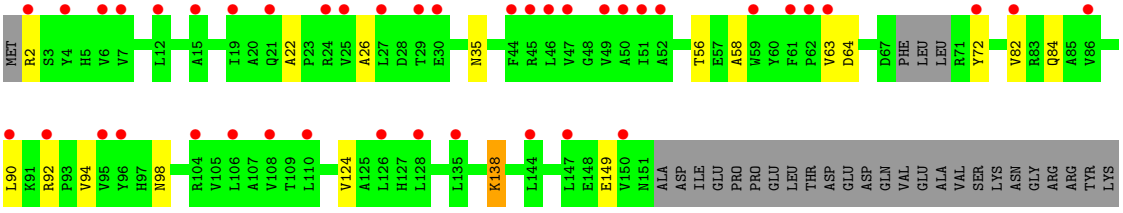
Chain A:

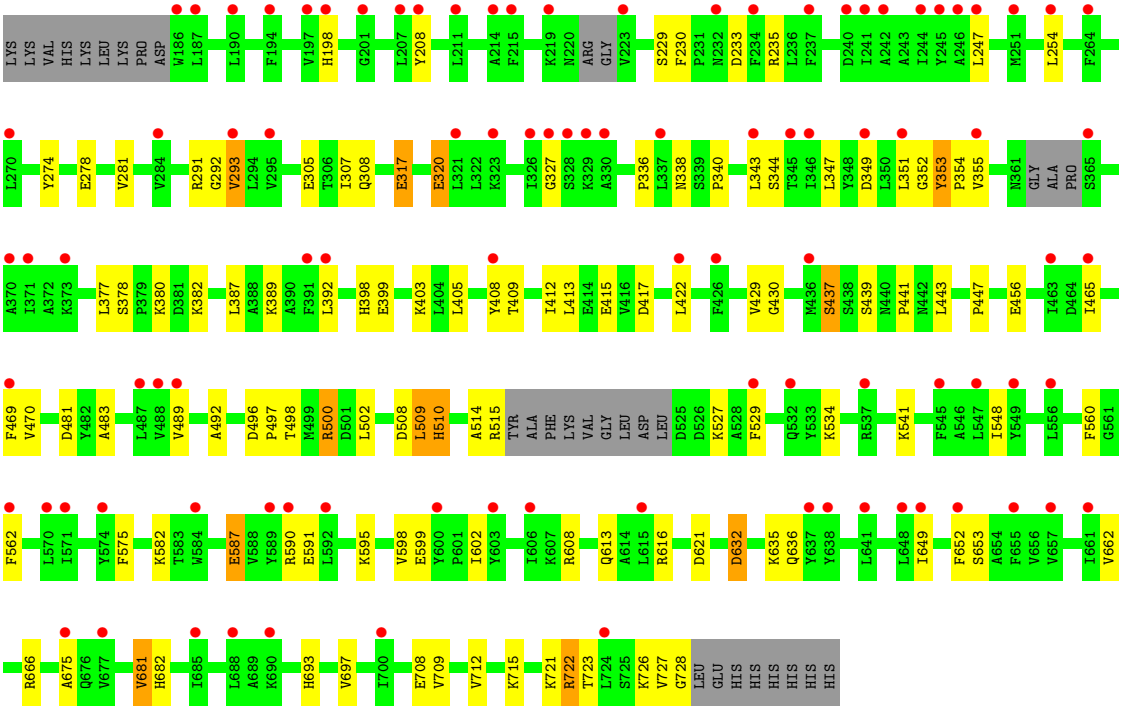




• Molecule 1: DNA polymerase I

Chain B:




LYS LYS VAL HIS LYS LEU LYS PRO ASP W186 L187 L190 F194 V197 H198 G201 L207 Y208 L211 A214 F215 K219 N220 ARG GLY V223 S229 F230 P231 N232 D233 F234 R235 L236 F237 D240 I241 A242 A243 I244 Y245 A246 L247 M251 L254 F264
L270 Y274 E278 V281 V284 R291 G292 V293 L294 V295 E305 T306 I307 Q308 E317 E320 L321 L322 K323 I326 G327 S328 K329 A330 P336 L337 N338 S339 P340 L343 S344 T345 I346 L347 Y348 D349 L350 L351 G352 Y353 P354 V355 N361 GLY ALA PRO S365
A370 I371 A372 K373 L377 S378 P379 K380 D381 K382 L387 A388 K389 A390 F391 L392 H398 E399 K403 L404 L405 Y408 T409 I412 L413 E414 E415 V416 D417 L422 F426 V429 G430 M436 S437 S438 S439 N440 P441 N442 L443 P447 E456 I463 D464 I465
F469 V470 D481 Y482 A483 L487 V488 V489 A492 D496 P497 T498 M499 R500 D501 L502 D508 L509 H510 A514 R515 TYR ALA PHE LYS VAL GLY LEU ASP LEU D525 D526 K527 A528 F529 Q532 Y533 K534 R537 K541 F545 A546 L547 I548 Y549 L556 F560 G561
F562 L570 I571 Y574 F575 K582 T583 W584 E587 V588 Y589 R590 E591 L592 K595 V598 E599 Y600 P601 I602 Y603 I606 K607 R608 Q613 A614 L615 R616 D621 D632 K635 Q636 Y637 Y638 L641 L648 I649 F652 S653 A654 F655 V656 V657 I661 V662
R666 A675 Q676 V677 V681 H682 I685 L688 A689 K690 H693 V697 I700 E708 V709 V712 K715 K721 R722 T723 L724 S725 K726 V727 G728 LEU GLU HIS HIS HIS HIS HIS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	309.77Å 98.01Å 77.61Å 90.00° 98.90° 90.00°	Depositor
Resolution (Å)	153.02 – 2.97 153.02 – 2.97	Depositor EDS
% Data completeness (in resolution range)	68.9 (153.02-2.97) 68.9 (153.02-2.97)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.59 (at 2.96Å)	Xtriage
Refinement program	BUSTER 2.11.8	Depositor
R, R_{free}	0.216 , 0.260 0.217 , 0.263	Depositor DCC
R_{free} test set	1564 reflections (4.79%)	wwPDB-VP
Wilson B-factor (Å^2)	107.0	Xtriage
Anisotropy	0.015	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	(Not available) , (Not available)	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	11067	wwPDB-VP
Average B, all atoms (Å^2)	124.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.80% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality

5.1 Standard geometry

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#$\|Z\|>5$	RMSZ	#$\|Z\|>5$
1	A	0.47	0/5785	0.63	2/7856 (0.0%)
1	B	0.41	0/5520	0.59	0/7488
All	All	0.44	0/11305	0.61	2/15344 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	448	ARG	N-CA-C	-5.37	96.50	111.00
1	A	470	VAL	N-CA-CB	-5.24	99.97	111.50

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry-related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	5660	0	5667	83	0
1	B	5407	0	5407	80	0
All	All	11067	0	11074	160	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 7.

The worst 5 of 160 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:447:PRO:HD2	1:B:465:ILE:HD13	1.58	0.83
1:A:489:VAL:HG21	1:A:653:SER:HB3	1.62	0.81
1:B:727:VAL:HG12	1:B:728:GLY:H	1.46	0.80
1:A:69:LEU:HD11	1:A:226:TYR:HB3	1.66	0.77
1:B:681:VAL:HG13	1:B:682:HIS:ND1	2.02	0.74

There are no symmetry-related clashes.

5.3 Torsion angles (i)

5.3.1 Protein backbone (i)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	704/736 (96%)	656 (93%)	44 (6%)	4 (1%)	25	61
1	B	664/736 (90%)	616 (93%)	40 (6%)	8 (1%)	13	45
All	All	1368/1472 (93%)	1272 (93%)	84 (6%)	12 (1%)	17	53

5 of 12 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	615	LEU
1	B	327	GLY
1	B	514	ALA
1	A	228	GLY
1	B	509	LEU

5.3.2 Protein sidechains (i)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	607/632 (96%)	574 (95%)	33 (5%)	22	55
1	B	580/632 (92%)	548 (94%)	32 (6%)	21	55
All	All	1187/1264 (94%)	1122 (94%)	65 (6%)	21	55

5 of 65 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	587	GLU
1	B	621	ASP
1	A	622	LYS
1	A	621	ASP
1	B	632	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	398	HIS
1	B	319	GLN
1	B	398	HIS

5.3.3 RNA (i)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

There are no ligands in this entry.

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

6 Fit of model and data

6.1 Protein, DNA and RNA chains

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95^{th} percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2			OWAB(Å^2)	Q<0.9
1	A	708/736 (96%)	0.90	82 (11%)	4	2	73, 106, 156, 165	0
1	B	676/736 (91%)	1.10	138 (20%)	1	0	92, 131, 190, 214	0
All	All	1384/1472 (94%)	1.00	220 (15%)	1	1	73, 118, 176, 214	0

The worst 5 of 220 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	194	PHE	7.4
1	B	7	VAL	6.8
1	B	245	TYR	6.7
1	B	27	LEU	6.3
1	B	190	LEU	6.1

6.2 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates

There are no monosaccharides in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.